**Do ecological differences between taxonomic groups influence the relationship between species' distributions and climate? A global meta-analysis using species distribution models**

Authors: Heather M. Kharouba[1*]

Jenny L. McCune[1]

Wilfried Thuiller[2]

Brian Huntley[3]

[1]The Biodiversity Research Centre, University of British Columbia, 2212 Main Mall, Vancouver, BC, V6T 1Z4, CANADA

[2] Laboratoire d'Ecologie Alpine, UMR-CNRS 5553, Université Joseph Fourier, BP53, 38041 Grenoble Cedex 9, France

[3] School of Biological and Biomedical Sciences, Durham University, South Road, Durham, DH1 3LE, United Kingdom

*Corresponding author: Heather M. Kharouba

kharouba@zoology.ubc.ca

**ABSTRACT**

Understanding whether and how ecological traits affect species' geographic distributions is a fundamental issue that bridges ecology and biogeography. While climate is thought to be the major determinant of species' distributions, there is considerable variation in the strength of species' climate-distribution relationships. One potential explanation is that species with relatively low dispersal ability cannot reach all geographic areas where climatic conditions are suitable. We tested the hypothesis that species from different taxonomic groups varied in their climate-distribution relationships because of differences in life history strategies, in particular dispersal ability. We conducted a meta-analysis by combining the discrimination ability (AUC values) from 4317 species distribution models (SDMs) using fit as an indication of the strength of the species' climate-distribution relationship. We found significant differences in the strength of species' climate-distribution relationships across taxonomic groups, however we did not find support for the dispersal hypothesis. Our results suggest that relevant ecological trait variation among broad taxonomic groups may be related to differences in species' climate-distribution relationships but which ecological traits are important remains unclear.

**Keywords**

Species distribution models, climate, taxonomic groups, dispersal, AUC, niche, ecological traits

## INTRODUCTION

Understanding whether and how ecological traits affect species' geographic distributions is a fundamental issue that bridges ecology and biogeography (Brown 1995, Wiens 2011). This issue has become even more relevant as ecologists and biogeographers struggle to understand the variation in species' responses to climatic change. For example, recent studies have examined the relationship between species' ecological traits, such as dispersal ability and ecological generalization, and changes in their distributions and phenology with recent climatic changes (Angert et al. 2011, Diamond et al. 2011). Identifying characteristics of organisms that determine their sensitivity to environmental change is crucial to ecological forecasting and conservation planning.

Central to this work is the theory of the niche: the set of abiotic and biotic conditions within which a species can persist (Hutchinson 1957). A species' distribution is limited to geographic areas where all these conditions meet the species' niche requirements. At broad spatial scales, climate has long been considered the most important factor in determining species' distribution limits (e.g. Merriam 1894, Good 1931, Gaston 2003). However, there seems to be considerable variation in the degree to which species' distributions are predicted by climate. There are three potential reasons for this variation. First, other abiotic or biotic factors may prevent a species from persisting even where the climate is suitable (Luoto et al. 2007). Alternatively, regions of suitable climate may be separated by areas that are not suitable which the species does not have sufficient dispersal ability to cross (Blach-Overgaard et al. 2010, Graham et al. 2010). Finally, if the species is relatively new and/or the climate has only recently become suitable, the species may not have had enough time to reach all suitable areas (Paul et al. 2009, Blach-Overgaard et al. 2010).

Dispersal ability is thought by some to determine how closely a species' current distribution matches the geographic distribution where all abiotic and biotic conditions meet its niche requirements. Species that produce many propagules that travel long distances are more likely to be able to cross any unsuitable habitat, and thus should be more likely to be found everywhere the climate is suitable. Therefore, dispersal ability may determine the strength of the species' climate-distribution relationship. Indeed, some studies have found evidence that dispersal ability can strongly affect species' distributions (e.g. Thuiller et al. 2004, Poyry et al. 2008). However, others suggest that the dispersal of individuals happens over such small time

scales relative to the formation of species' geographic distributions that it has little importance (Lester et al. 2007).

Many have hypothesized that species in different taxonomic groups should vary in their species' climate-distribution relationship because of their different life history strategies, in particular dispersal ability (e.g. Araújo and Pearson 2005, Wisz et al. 2008). The fit of species distribution models (SDMs) has often been used to test this hypothesis (Araújo and Pearson 2005, Tsoar et al. 2007). SDMs use various statistical techniques to describe the relationship between observed environmental variables, such as mean annual temperature, and the recorded spatial occurrence (presence/absence) of a species (see e.g. Guisan and Zimmermann 2000). The ability of an SDM based only on climatic factors to predict the presence or absence of a species can be considered an indication of the strength of the species' climate-distribution relationship: the greater the success of a SDM at predicting the species' presence/absence in a given location, the stronger the correlation between climatic variables and the presence/absence of the species. Some studies have found species' climate-distribution relationship differences between taxonomic groups (Araújo and Pearson 2005, Tsoar et al. 2007), whereas others have not (Pearce and Ferrier 2000, Wisz et al. 2008). It is unclear whether these varying results are due to the different geographic regions, groupings of species, or modeling techniques of each study. Despite the availability of SDMs for thousands of species, a comprehensive comparison of the fit of SDMs between different taxonomic groups has not been made.

Here, we tested the hypothesis that taxonomic groups varied in the strength of their species' climate-distribution relationships. We predicted that taxonomic groups with lower dispersal ability would have weaker species' climate-distribution relationships. We used a meta-analysis approach and combined the discrimination ability metrics that were reported from 4317 SDMs in twenty studies using only climatic variables to determine whether species varied predictably in their climate-distribution relationships based on taxonomic affinities. We also compiled dispersal distances for a subset of these species to determine whether dispersal ability *directly* influenced the strength of species' climate-distribution relationships. To facilitate a quantitative comparison we used a standardized discrimination ability measure and accounted statistically for methodological differences among studies.

**MATERIALS AND METHODS**

**Comment [HK1]:** We removed 'all else being equal' as we no longer think it is necessary to include here. We had initially meant this to refer to methodological considerations such as the variables chosen are the most important etc. but we no longer think this level of detail is needed here.

**Data compilation**

We conducted a literature search using Web of Science for studies (published before March 2009) that reported statistical measures of goodness-of-fit for SDMs constructed for individual species based on climatic variables only. We searched for studies using the terms "ecological niche model" and "climat*", "species distribution model" and "climat*", and "climate envelope model" and "climat*. Studies were *excluded* if: (1) one or more non-climatic variables, such as soil fertility, land use or land cover, were included in the SDM; (2) model fit was measured only qualitatively or not reported; or (3) model fit was reported only as averages across species. In cases where model fit was not reported for all individual species modeled, we requested these data from the authors. Due to the small number of studies modeling aquatic species, we limited our analysis to terrestrial species.

We needed a metric of model fit that was comparable across studies. We found AUC (area under a receiver operating characteristic curve) to be the most common metric (other metrics: Cohen's kappa, sensitivity, specificity, range filling rates), therefore our analysis was limited to studies that reported AUC. AUC measures the ability of a SDM to discriminate sites where a species is present from sites where it is absent, rather than goodness-of-fit *per se*. It considers the relationship between false-positives and true-positives and ranges from zero to one, where perfect discrimination gives a value of one (Fielding and Bell 1997). Hereafter, we use the term SDM 'fit' to indicate 'discrimination ability' as measured by AUC. When studies reported AUC for both training and test data, test AUC values were used. Although this metric has been criticised (e.g. Lobo et al. 2008), it was the only measure in common across most of the studies.

Some species' distributions were modeled several times, either by the same study (using multiple modeling techniques (n=9) or resolutions (n=1)) or by several studies (most such species were modeled by only two studies). In all cases, we randomly selected one SDM per species and used the associated AUC value and methodology. This produced a dataset of 4317 species and their SDMs from twenty studies (Supplementary material Appendix 1-3). These studies modeled species in Europe (10 studies, 2301 spp.), North America (2 studies, 67 spp.), South America (2 studies, 32 species) and Africa (6 studies, 1917 spp.). We classified each species into one of five broad taxonomic groups: mammals (483 spp.), butterflies (116 spp.), herptiles (reptiles and amphibians; 114 spp.), birds (2099 spp.), and plants (1505 spp.).

**Comment [HK2]:** We included this list in response to a comment by one of the reviewers that said we should at least mention the other possible metrics. We are happy to remove it if it is no longer thought to be necessary.

134 SDM fit can be affected by the type of model used (e.g. Elith et al. 2006), the number of
135 climatic variables used (e.g. Pearce and Ferrier 2000), the resolution or grain size used (e.g.
136 Guisan et al. 2007), the total extent over which the species' range was modeled (e.g. Luoto et al.
137 2005), and latitude (Brown et al. 1996, Luoto et al. 2005). Therefore, for each SDM we noted
138 the modeling technique, number of distinct climatic variables used in the model, resolution
139 ($km^2$), total spatial extent ($km^2$) and average absolute latitude and then included these as
140 covariates in our statistical analysis.
141 Another factor which may lead to differences in SDM fit between species is prevalence
142 (McPherson et al. 2004, Santika, 2011), the number of grid cells from which a species is
143 recorded as present expressed as a proportion of the total number of grid cells from which data
144 are available. We were able to obtain prevalence values for almost all of the SDMs (n=4089),
145 allowing us to explore any effects of prevalence on SDM fit.
146 Finally, we scanned the literature to find dispersal distances for as many of our species as
147 possible to assess whether there were significant differences in measured dispersal ability among
148 our taxonomic groups. True dispersal distances are very difficult to measure due to phenomena
149 such as very rare long-distance dispersal events. Therefore, we used the directly measured ability
150 of an organism or its propagules to move (i.e. its mobility) as an estimate of a species' dispersal
151 distance. We considered both maximum and mean measured dispersal distances but excluded
152 migratory distances to standardize measures of dispersal distances across taxonomic groups.
153 Where more than one distance was reported per species or study we used the mean of mean
154 distances, and the maximum of maximum distances. We found mean dispersal distances for 241
155 species for which we also had AUC values (birds=103, butterflies=22, mammals=22, plants=94).
156 For maximum dispersal distance, we found 105 species that also had AUC values (birds=27,
157 butterflies=18, mammals=30, plants=18). For further details, see Supplementary material
158 Appendix 4,5.
159
160 **Statistical analysis**
161 There were two parts to the analysis. The first was to determine whether there were any
162 significant differences in SDM fit between taxonomic groups and whether those differences were
163 robust to potential confounding factors (covariates). The second was to explore the relationship
164 between SDM fit, taxonomic group and the other covariates. We used generalized linear mixed-

effects models (GLMM, glmmadmb function in the "glmmADMB" package (Skaug et al. 2012) in R (R Development Core Team 2012)) with a Beta error distribution with AUC as our response variable and 'study' as a random factor. AUC values of exactly one, which are not allowed with the beta distribution, were converted to 0.99 instead (eight significant digits were used to ensure a unique value and to match the maximum precision of the data, n=117). To allow for model estimation, we collapsed the six rarest modeling types into one category to reduce the number of types (from 18 to 12; these six techniques were used for only 0.35% of all SDMs). We took the logarithm of spatial extent to improve normality (except in the collinearity test), but all other covariates were used without transformation. Taxonomic group and model type were categorical, and all other covariates were continuous.

*Relationship between discrimination ability and taxonomic group*

To test whether taxonomic group explained significantly more deviance in AUC than expected at random, we compared a model with only an intercept to a model with only taxonomic group. We then tested whether differences in discrimination ability across taxonomic groups explained significant additional deviance after accounting for the combined effect of the differences in the methodological approach of studies (i.e. the covariates: model type, resolution, number of climatic variables, spatial extent and latitude). For all model comparisons, we used a likelihood ratio test. We also calculated AIC for all models to evaluate the relative effects of individual covariates.

We first inspected bivariate plots of all continuous covariates before constructing pairwise correlations to identify potential problems with multi-collinearity among covariates (Supplementary material Appendix 6). Latitude was highly correlated with spatial extent and resolution (Spearman's r = - 0.903, -0.589 respectively, n = 4317, Supplementary material Appendix 6) and explained less deviance in AUC than spatial extent or resolution (Table 1), therefore the 'full model' included taxonomic group, model type, spatial extent and number of climatic variables. We considered the effect of 'study' by including it as a random factor and by testing the influence of individual studies that contributed more than half of the total number of species in one taxonomic group ("large studies") by comparing results obtained with and without each of these studies (Huntley et al. 2006, Araújo et al. 2005, Luoto et al. 2005, Supplementary material Appendix 3).

196

197 *Relationship between SDM fit, covariates and taxonomic group*

198 We tested whether individual covariates (including prevalence) explained significantly more
199 deviance in AUC than under random expectation and after accounting for all other covariates
200 (including taxonomic group) by comparing each model to a reduced one. Finally, to test whether
201 there were significant differences in dispersal distance (both mean and maximum) across
202 taxonomic groups, we used a Kruskal-Wallis rank sum test. We then tested whether dispersal
203 distance explained significantly more deviance in AUC by comparing a model with and without
204 dispersal distance. Dispersal distance was log-transformed to improve normality. Lastly, to test
205 for the possibility that an interaction between dispersal distance and taxonomic group explained
206 deviance in AUC, we compared a model with and without this two-way interaction.

207 All statistical analyses were performed using R 2.14.1 (R Development Core Team 2012).

208

209 **RESULTS**

210 **Relationship between discrimination ability and taxonomic group**

211 Mean AUC across all species was 0.941 (±0.00104 SE, n=4317). Birds had the highest
212 mean AUC (0.954 ±0.00145 SE, n=2099) and butterflies had the lowest mean AUC (0.856
213 ±0.0114 SE, n=116; Fig. 1a). However, the ranking and pair-wise comparison of taxonomic
214 groups changed depending on which "large study" was removed (Fig. 1).

215 Taxonomic group explained significant deviance in AUC ($LRT_{7,3}$=46.98, $p$<0.0001;
216 Table 1), even after accounting for all covariates ($LRT_{20,16}$=46.64, $p$<0.0001; Table 1). The
217 effect of taxonomic group was also robust to the exclusion of each of the "large studies"
218 (Supplementary material Appendix 7).

219

220 **Relationship between discrimination ability, covariates and taxonomic group**

221 SDM model type explained significant deviance in AUC ($LRT_{3,13}$=120.58, $p$<0.0001;
222 Table 1), even after accounting for all the other covariates ($LRT_{20,10}$=120.14, $p$<0.0001; Table 1).
223 For the subset of species for which we had prevalence data, prevalence also explained significant
224 deviance in AUC after accounting for all covariates (including taxonomic group;
225 $LRT_{12,11}$=447.62, $p$<0.0001; Table 1). SDMs with greater prevalence had lower AUC
226 (Spearman's r= -0.4937).

In our subset of species with dispersal distances, mean dispersal distance was greatest for mammals (175 km) while birds had the greatest maximum dispersal distance (1305 km; Fig. 2). Butterflies had the shortest mean and maximum dispersal distance (0.441 km and 2.25 km, respectively; Fig. 2). The ranking of groups closely matched the ranking of groups of the entire dataset in terms of AUC for both dispersal measures (Fig. 1a, Fig.2). There was also a significant difference between taxonomic groups in dispersal distance (mean: df=3, $\chi^2$=181.006, $p$<0.0001; max: df=4, $\chi^2$=291.557, $p$<0.0001). Taxonomic group explained significant deviance in AUC (mean: $LRT_{6,3}$=10.386, p=0.01555; max: $LRT_{7,3}$=13.022, $p$=0.01117). However, dispersal distance did not explain significant deviance in AUC (mean: $LRT_{4,3}$=2.068, $p$=0.1504; max: $LRT_{4,3}$=0.144, $p$=0.7043). There was no significant interaction between taxonomic group and dispersal distance (mean: $LRT_{10,7}$=4.508, $p$=0.2116; max: $LRT_{12,8}$=4.506, $p$=0.3418).

## DISCUSSION

We found support for taxonomic differences in SDM fit suggesting a role for ecological traits in affecting species' geographic distributions at broad scales. However, prevalence and methodological issues, such as model type, also influenced SDM fit. Indeed, both factors have been shown previously to influence SDM fit (e.g. Elith et al. 2006, Santika 2011). We also found that "large studies" influenced the relationship among taxonomic groups and AUC, for example the taxonomic group with the highest mean AUC varied with the subset of species considered (Fig. 1). Therefore, species' taxonomic affinities, prevalence and methodological issues, such as the model type, are all important in influencing species' climate-distribution relationships as measured by SDMs.

There are a number of potential explanations for the difference in the strength of species' climate-distribution relationships between taxonomic groups. First, taxonomic differences may reflect differences in dispersal ability among groups. Certainly, we found differences in measured dispersal distances between broad taxonomic groups that were consistent with the dispersal hypothesis (Fig. 1a, Fig. 2). However, there were inconsistencies in the ranking and pair-wise comparisons of taxonomic groups in SDM fit depending on the subset of species considered (Fig. 1). Moreover, there was no significant relationship between AUC and dispersal distance. Therefore, our results indicate that greater dispersal ability, at least in terms of measurable differences in mobility, may not result in stronger overall species' climate-

distribution relationships at broad scales. However, dispersal distance is inherently difficult to measure and our estimate of dispersal ability may not have been the most appropriate for all species. For example, we did not take into account migratory or rare long-distance dispersal events. Consequently, we may have underestimated the role of dispersal ability for certain species.

Alternatively, dispersal may not be an important trait in determining species' climate-distribution relationships. The majority of species had low prevalence (77% species had <0.1 prevalence) and species with lower prevalence were more likely to have higher AUC values. If these low prevalence species are mainly specialists (i.e. restricted range endemics) that are adapted to uncommon climatic conditions found in small, contiguous areas, they could have strong climate-distribution relationships regardless of dispersal ability.

Third, other life history traits, for example, body size, generation time or diet breadth, may influence the strength of species' climate-distribution relationships between taxonomic groups. However, determining their relative importance may be difficult across the broad taxonomic groups considered. Lower-order taxonomic groups, or functional groups of species within or across taxonomic groups, might be more effective in dividing species according to relevant traits. Nevertheless, while some recent studies dividing species into finer taxonomic or functional divisions have found significant differences in species' climate-distribution relationships (e.g. Syphard and Franklin 2009), others have not (e.g. Huntley et al. 2004).

On the other hand, taxonomic differences in SDM fit may be a function of the sample unbalance (across studies and taxonomic groups; Supplementary material Appendix 2) and the high average discrimination ability. Both of these factors could reflect issues related to fitting, testing and publishing SDMs. SDMs have been criticized for not using independent data to test their models (e.g. Hampe 2004, Segurado et al. 2006). Without independent test occurrence points, well-fitting models could reflect spatial autocorrelation between training and testing points rather than relationships between species' presence/absence and climatic variables. Moreover, SDMs may be overfitted by fitting complex response curves and re-fitting models until a high AUC is achieved (Araújo et al. 2005, Guisan and Thuiller 2005). We also suggest that there could be a "file-drawer" problem, whereby species that do not achieve a high enough AUC value based on the literature standard (Swets 1988) are not published. In particular, when the objective of fitting the SDM is to predict species' potential distribution shifts under various

289 climate change scenarios, authors (rightly) do not use SDMs with very low discrimination
290 ability. For example, of the 453 species that Huntley et al. (2008) modeled, 13 native species that
291 did not yield "useful" models (*sensu* Swets 1988) were excluded from the synthesis. Taken
292 together, these issues could inflate AUC values and reduce overall variation, making it difficult
293 to detect the true relationship between taxonomic groups. While we acknowledge these
294 limitations of SDMs, to our knowledge, there are no other comparable published metrics to
295 evaluate individual species' climate-distribution relationships at such large scales. Moreover,
296 SDMs are still being used to better understand the relationship between species' distributions and
297 climate (e.g. Blach-Overgaard et al. 2010, Graham et al. 2010).
298 Lastly, because SDMs are fitted to species' current distributions they reflect both direct and
299 indirect influences of climate on those distributions. Non-climatic factors that limit a species to
300 certain broad areas (such as biotic interactions or other abiotic factors) are generally modulated
301 by climatic conditions. For example, since its introduction to Hawaii, avian malaria now
302 restricts native bird species to higher elevations, where temperature halts development of the
303 malaria pathogen inside its mosquito vector (van Riper et al. 1986). Differences among
304 taxonomic groups in the ability of climate to *directly* limit species' distributions thus cannot be
305 revealed by our data, given that the SDMs we used cannot differentiate direct from indirect
306 climatic effects. However, we have no *a priori* reason to expect cases where climate acts
307 principally indirectly to occur more frequently in one taxonomic group than another. In addition,
308 even if a species' distribution is indirectly limited by climate due to the climatic tolerances of a
309 competitor, predator, or disease, at broad scales, climate is still the ultimate determinant of the
310 species' distribution.
311 There are a number of steps to be taken in the future to clarify how ecological traits
312 influence species' climate-distribution relationships. Firstly, more SDMs are needed for some
313 taxonomic groups, particularly invertebrates and herptiles. Secondly, we should strive to
314 eliminate issues related to species distribution modeling by using spatially/temporally
315 independent training and test datasets where possible (e.g. Beerling et al. 1995, Randin et al.
316 2006). Third, analyzing SDM prediction errors might help to shed light on the mechanism
317 driving the variation in species' climate-distribution relationships, especially in cases of poor fit
318 (e.g. Hanspach et al. 2011). For example, SDMs with more false negatives overall than false
319 positives could suggest that source-sink dynamics are important: even where conditions are not

favourable, individuals may still persist owing to a rescue effect, or temporal variation in conditions (Gaston 2003, Pulliam 2000). Alternatively, models with greater rates of false positives might suggest that dispersal limitation or interspecific interactions, such as competition, are limiting a species' distribution (Pulliam 2000, Graham et al. 2010). Finally, exploring spatial variation in model behaviour, for example testing model performance in climatically heterogeneous regions or through patterns of spatial prediction errors (Hanspach et al. 2011), could also improve our understanding of model performance and thus species' climate-distribution relationships.

## CONCLUSION

We found a statistically significant effect of membership in broad taxonomic groups on SDM fit even after accounting for methodological issues, suggesting a role for ecological traits in determining the strength of species' climate-distribution relationships. However, the study itself, the model type used to build the SDM and species' prevalence all had significant effects on discrimination ability. Our results did not the support the hypothesis that dispersal ability affects the strength of species' climate-distribution relationships. However, more work is needed to determine which ecological traits are important in determining the strength of this relationship, and at what spatial scale and taxonomic level they are manifested.

## ACKNOWLEDGEMENTS

Catherine Graham and three anonymous reviewers gave valuable comments on the manuscript. Thanks to O. Allouche, M. Bakkenes, J. Elith, R. Kadmon, J. McPherson, R. Pearson, J. Pöyry, and A. Tsoar for contributing supplemental data. M. Vellend, D. Srivastava, and J. Goheen provided helpful comments on previous versions of the manuscript and D. Schluter, B. Bolker, and D. Fournier gave statistical advice. This research was funded by Natural Sciences and Engineering Research Council of Canada (NSERC) through graduate scholarships to HMK and JLM. WT acknowledges support from the European Commission FP6 ECOCHANGE project (Grand GOCE-CT-2007-036866).

## REFERENCES

350 Angert, A.L. et al. 2011. Do species' traits predict recent shifts at expanding range edges? - Ecol.
351 Lett. 14: 677-689.
352 Araújo, M. B. and Pearson, R. G. 2005. Equilibrium of species' distributions with climate. -
353 Ecography 28: 693-695.
354 Araújo, M. B. et al. 2005. Validation of species-climate impact models under climate change. -
355 Global Change Biol. 11:1504-1513.
356 Beerling, D.J. et al. 1995. Climate and the distribution of *Fallopia japonica*: use of an introduced
357 species to test the predictive capacity of response surfaces. - J. Veg. Sci. 6:269-282.
358 Blach-Overgaard, A. et al. 2010. Determinants of palm species distributions across Africa: the
359 relative roles of climate, non-climatic environmental factors, and spatial constraints
360 33:380-391.
361 Brown, J.H. 1995. Macroecology. - The University of Chicago Press, Chicago, US.
362 Brown, J. H. et al. 1996. The geographic range: Size, shape, boundaries, and internal structure. -
363 Annu. Rev. Ecol., Evol. Syst. 27: 597-623.
364 Diamond, S.E. et al. 2011. Species' traits predict phenological responses to climate change in
365 butterflies. - Ecology 92: 1005-1012
366 Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence
367 data. - Ecography 29: 129-151.
368 Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors
369 in conservation presence/absence models. - Environ. Conserv. 24: 38-49.
370 Gaston, K. J. 2003. The structure and dynamics of geographic ranges, 1st edn. - Oxford
371 University Press, Oxford.
372 Graham, C. H., et al. 2010. Evaluating the potential causes of range limits of birds of the
373 Colombian Andes. J. Biogeogr. 37: 1863-1875.
374 Good, R. D. O. 1931. A theory of plant geography. - New Phytol. 30: 149-149.
375 Guisan, A. et al. 2007. Sensitivity of predictive species distribution models to change in grain
376 size. - Divers. Distrib. 13: 332-340.
377 Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple
378 habitat models. - Ecol. Lett. 8: 993-1009.
379 Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. -
380 Ecol. Model. 135:147-186.

381 Hampe, A. 2004. Bioclimatic models: what they detect and what they hide. - Global Ecol.
382 Biogeogr. 11: 469–471.
383 Hanspach et al. 2011. Geographical patterns in prediction errors of species distribution models. -
384 Global Ecol. Biogeogr. 20: 779-788.
385 Huntley, B. et al. 2004. The performance of models relating species geographical distributions to
386 climate is independent of trophic level. - Ecol. Lett. 7: 417-426.
387 Huntley, B. et al. 2006. Potential impacts of climatic change upon geographical distributions of
388 birds. - Ibis 148: 8-28.
389 Huntley, B. et al. 2008. Potential impacts of climatic change on European breeding birds. - PLoS
390 ONE 3: e1439.
391 Hutchinson, G. E. 1957. Concluding Remarks. - Cold Spring Harbor Symposium on Quantitative
392 Biology 20: 415-427.
393 Lester, S. E. et al. 2007. The relationship between dispersal ability and geographic range size. -
394 Ecol. Lett. 10: 745-758.
395 Lobo, J. M. et al. 2008. AUC: a misleading measure of the performance of predictive distribution
396 models. - Global Ecol. Biogeogr. 17: 145-151.
397 Luoto, M. et al. 2005. Uncertainty of bioclimate envelope models based on the geographical
398 distribution of species. - Global Ecol. Biogeogr. 14: 575-584.
399 Luoto, M. et al. 2007. The role of land cover in bioclimatic models depends on spatial resolution.
400 - Global Ecol. Biogeogr. 16: 34-42.
401 McPherson, J. M. et al. 2004. The effects of species' range sizes on the accuracy of distribution
402 models: ecological phenomenon or statistical artefact? - J. Appl. Ecol. 41: 811-823.
403 Merriam, C. H. 1894. Laws of temperature control of the geographic distribution of terrestrial
404 animals and plants. - National Geographic 6: 229-238.
405 Paul, J.R., Morton, C., Taylor, C. M. & Tonsor, S. J. (2009) Evolutionary time for dispersal
406 limits the extent but not the occupancy of species' potential ranges in the tropical plant
407 genus Psychotria (Rubiaceae). Am. Nat. 173:188-199.
408 Pearce, J. and Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species
409 distribution models using logistic regression. - Ecol. Model. 128: 127-147.
410 Poyry, J. et al. 2008. Species traits are associated with the quality of bioclimatic models. - Global
411 Ecol. Biogeogr. 17: 403-414.

412 Pulliam, H. R. 2000. On the relationship between niche and distribution. - Ecol. Lett. 3: 349-361.
413 R Development Core Team. 2012. R: A Language and Environment for Statistical Computing.
414 R Foundation for Statistical Computing,Vienna, Austria. http://www.R-project.org
415 Randin, C.F. et al. 2006. Are niche-based species distribution models transferable in space? - J.
416 Biogeogr. 33: 1689-1703.
417 Santika, T. 2011. Assessing the effect of prevalence on the predictive performance of species
418 distribution models using simulated data. - Global Ecol. Biogeogr. 20: 181-192.
419 Segurado, P. et al. 2006. Consequences of spatial autocorrelation for niche-based models. - J.
420 Appl. Ecol. 43: 433-444.
421 Skaug, H. et al. 2012. glmmADMB: Generalized Linear Mixed Models Using AD Model
422 Builder. - R package version 0.7.2.12.
423 Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. - Science 240: 1285-1293.
424 Syphard, A. D. and Franklin, J. 2010. Species traits affect the performance of species distribution
425 models for plants in southern California. - J. of Veg. Sci. 21: 177-189.
426 Thuiller, W. et al. 2004. Relating plant traits and species distributions along bioclimatic gradients
427 for 88 Leucadendron taxa. - Ecology 85: 1688-1699.
428 Tsoar, A. et al. 2007. A comparative evaluation of presence-only methods for modeling species
429 distribution. - Divers. Distrib. 13: 397-405.
430 van Riper et al. 1986. The epizootiology and ecological significance of malaria in Hawaiian land
431 birds. - Ecol. Monogr. 56: 327-344.
432 Wiens, J. J. 2011. The niche, large-scale biogeography, and species interactions. - Philos. Trans.
433 R. Soc. Lond., Ser. B: Biol. Sci. 366: 2336–2350.
434 Wisz, M. S. et al. 2008. Effects of sample size on the performance of species distribution models.
435 - Divers. Distrib. 14: 763-773.
436
437

**TABLE Legends**

Table 1. Analysis of deviance table for the relationship between discrimination ability, covariates and taxonomic group. Presented are the differences in degrees of freedom, AIC and deviance between full and reduced models as well as the associated *p* value. Models are compared for all species (n=4317) and for the subset of species with prevalence values (n=4089). Depending on the model comparison and term of interest, the full model includes all other covariates (number of variables, log(spatial extent), model type, resolution and taxonomic group).

447 Table 1.

| Model for comparison | Data | Model terms | Difference in d.f. | ΔAIC | ΔDeviance | *p* |
|---|---|---|---|---|---|---|
| Just intercept | All species | Intercept | | | | |
| | | + taxonomic group | 4 | 38.98 | 46.98 | <0.0001 |
| | | + model type | 10 | 100.58 | 120.58 | <0.0001 |
| | | + log(spatial extent) | 1 | 2.58 | 4.58 | 0.03235 |
| | | + resolution | 1 | 1.38 | 3.38 | 0.0660 |
| | | + number of climatic variables | 1 | 1.20 | 0.8 | 0.3711 |
| | | + latitude | 1 | 0.58 | 1.42 | 0.2334 |
| | Subset | Intercept | | | | |
| | | + prevalence | 1 | 335.36 | 337.36 | <0.0001 |
| Full model | All species | Full model | | | | |
| | | + taxonomic group | 4 | 38.64 | 46.64 | <0.0001 |
| | | + model type | 10 | 101.52 | 120.14 | <0.0001 |
| | | + log(spatial extent) | 1 | 1.12 | 3.12 | 0.0773 |
| | | + resolution | 1 | -1.38 | 0.62 | 0.431 |
| | | + number of climatic variables[α] | NA | NA | NA | NA |
| | Subset | Full model | | | | |
| | | + prevalence[†] | 1 | 445.62 | 447.62 | <0.0001 |

448 [α] No solution was found

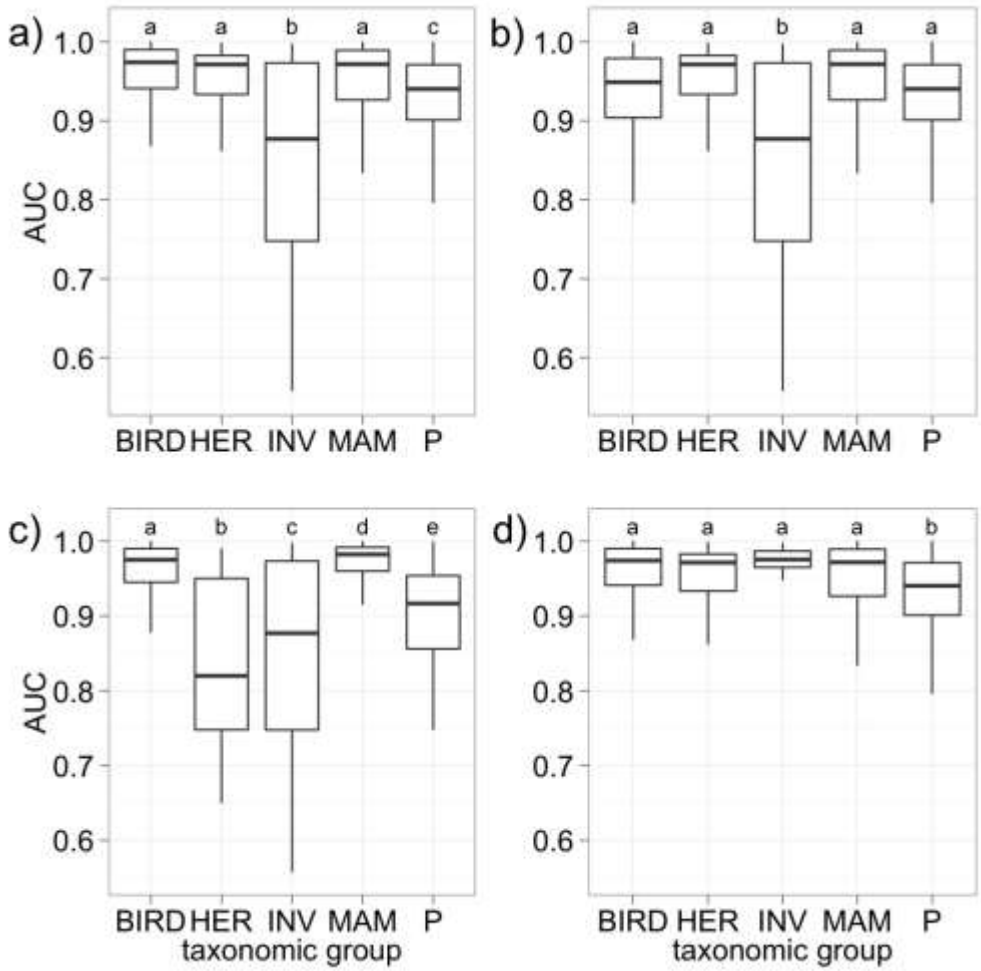
449 [†]A model solution could only be found if number of climatic variables was not included

**FIGURE LEGENDS**

**Figure 1.** Taxonomic differences in discrimination ability (AUC) across all studies (based on 4317 species from twenty published studies (number of species: birds n=2099; herptiles n=114; butterflies n=116; mammals n=483; plants n=1505)) (a), without Huntley et al. 2006 (based on 2860 species from nineteen published studies (number of species: birds n=642; herptiles n=114; butterflies n=116; mammals n=483; plants n=1505)) (b), without Araújo et al. 2005 (based on 2539 species from nineteen published studies (number of species: birds n=1942; herptiles n=11; butterflies n=116; mammals n=331; plants n=139)) (c), and without Luoto et al. 2005 (based on 4238 species from nineteen published studies (number of species: birds n=2099; herptiles n=114; butterflies n=37; mammals n=483; plants n=1505)) (d). Taxonomic groups represented are: "BIRD"= birds, "HER"= herptiles, "INV"= butterflies, "MAM"= mammals, "P"= plants. Taxonomic groups with different letters above them are significantly different according to pair-wise comparisons. Outliers were removed to improve visual contrasts between taxonomic groups.

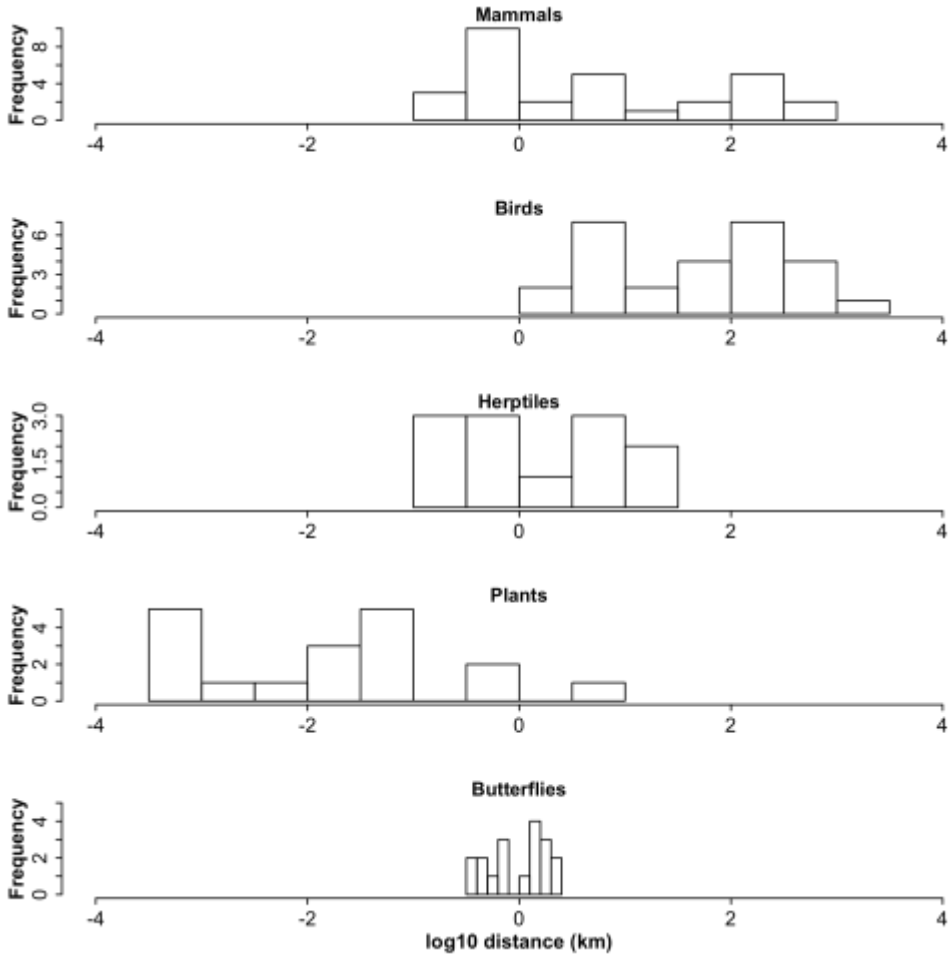
**Figure 2**. Taxonomic differences in log (base 10) maximum dispersal distances (km) for 105 species (birds=27, butterflies=18, mammals=30, plants=18).

468 Figure 1.




469

Figure 2.


Mammals
Frequency
0 4 8
-4 -2 0 2 4
Birds
Frequency
0 3 6
-4 -2 0 2 4
Herptiles
Frequency
0.0 1.5 3.0
-4 -2 0 2 4
Plants
Frequency
0 2 4
-4 -2 0 2 4
Butterflies
Frequency
0 2 4
-4 -2 0 2 4
log10 distance (km)

476 **SUPPLEMENTARY MATERIAL**

477

478 **Appendix 1** Methodological attributes used to build SDMs for each study used in the
479 analysis.

480 **Appendix 2** Taxonomic attributes of studies used in the analysis.

481 **Appendix 3** Full references for studies cited in Appendix 1 and 2.

482 **Appendix 4** Attributes of studies that contained dispersal distances for species in our dataset.

483 **Appendix 5** Full reference for studies cited in Appendix 4.

484 **Appendix 6** Collinearity (Spearman r coefficients) between all continuous covariates
485 (n=4317).

486 **Appendix 7** Analysis of deviance table for the relationship between model accuracy,
487 covariates and taxonomic group when studies that contributed more than half of the total
488 number of species in one taxonomic group were removed.

489 **Appendix 1.** Methodological attributes used to build SDMs for each study used in the analysis. Presented are the model types, number of
490 climatic variables used in the model, resolution ($km^2$) of the model, total spatial extent over which the model was built ($km^2$) and average
491 absolute latitude of the region for which the model was built.

| Study | Model type (s) | Variables | Resolution ($km^2$) | Spatial extent ($km^2$) | Latitude ($^o$) |
|---|---|---|---|---|---|
| Araújo et al. 2005 | GAM | 7 | 2500 | $1.105x10^7$ | 47 |
| Huntley et al. 2006 | GAM, locally weighted regression | 4 | 12227 | $2.40x10^7$ | 0 |
| Huntley et al. 2008 | locally weighted regression | 3 | 2500 | $1.105x10^7$ | 47 |
| Huntley et al. 2004 | locally weighted regression | 3 | 2500 | $1.105x10^7$ | 47 |
| Beale et al. 2008 | ANN | 3 | 2500 | $6.04x10^6$ | 47 |
| Thuiller et al. 2006 | GAM | 6 | 256 | $3.02x10^7$ | 0 |
| McPherson and Jetz 2007 | autologistic regression | 1-28 | 2975 | $8.27x10^6$ | 47 |
| Elith et al. 2006 | Mars, gdm, maxent, brt, domain, bruito, GAM, GARP, GLM, bioclim, lives | 11 | 1 | $1.465 x10^7$ | 14 |
| Heikkinen et al. 2007 | GAM | 3 | 100,1600 | $3.381 x10^5$ | 64 |

| | | | | | |
|---|---|---|---|---|---|
| Luoto et al. 2005 | GAM | 3 | 100 | $3.381 \times 10^5$ | 47 |
| Parra and Monahan 2008 | maxent | 19 | 16 | $4.240 \times 10^5$ | 15 |
| Phillips et al. 2006 | GARP, maxent | 13 | 30.25 | 19621904 | 15 |
| Thuiller 2003 | GLM, CART | 7 | 2500 | 5222500 | 15 |
| Freedman et al. 2009 | maxent | 7 | 1 | 475442 | 6 |
| Guisan and Hofer 2003 | GLM | 12 | 1 | $4.10 \times 10^4$ | 47 |
| Venier et al. 2004 | logistic regression | 10 | 25 | $8.0 \times 10^5$ | 15 |
| Pearson et al. 2006 | ANN, GARP, GAM, CGM | 5 | 2.56 | $1.22 \times 10^6$ | 15 |
| McPherson et al. 2004 | logistic regression, discriminant | 61 | 648 | $2.77 \times 10^6$ | 15 |
| Thuiller et al. 2004 | GAM | 4 | 2500 | $6.525 \times 10^6$ | 15 |
| Thuiller et al. 2003 | GLM | 7 | 2500 | $1.105 \times 10^7$ | 15 |

**Appendix 2.** Taxonomic attributes of studies used in the analysis. Presented are the total number of unique species used, the number of birds, herptiles, butterflies, mammals, and plants.

| Study | Total number of species | Birds | Herptiles | Butterflies | Mammals | Plants |
|---|---|---|---|---|---|---|
| Araújo et al. 2005 | 1778 | 157 | 103 | 0 | 152 | 1366 |
| Huntley et al. 2006 | 1457 | 1457 | 0 | 0 | 0 | 0 |
| Huntley et al. 2008 | 214 | 214 | 0 | 0 | 0 | 0 |
| Huntley et al. 2004 | 173 | 36 | 0 | 37 | 0 | 100 |
| Beale et al. 2008 | 42 | 42 | 0 | 0 | 0 | 0 |
| Thuiller et al. 2006 | 272 | 0 | 0 | 0 | 272 | 0 |
| McPherson and Jetz 2007 | 176 | 176 | 0 | 0 | 0 | 0 |
| Elith et al. 2006 | 30 | 0 | 0 | 0 | 0 | 30 |
| Heikkinen et al. 2007 | 2 | 2 | 0 | 0 | 0 | 0 |
| Luoto et al. 2005 | 79 | 0 | 0 | 79 | 0 | 0 |
| Parra and Monahan 2008 | 57 | 0 | 0 | 0 | 57 | 0 |
| Phillips et al. 2006 | 2 | 0 | 0 | 0 | 2 | 0 |
| Thuiller 2003 | 2 | 0 | 0 | 0 | 0 | 2 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Freedman et al. 2009 | 3 | 0 | 3 | 0 | 0 | 0 |
| Guisan and Hofer 2003 | 8 | 0 | 8 | 0 | 0 | 0 |
| Venier et al. 2004 | 10 | 10 | 0 | 0 | 0 | 0 |
| Pearson et al. 2006 | 4 | 0 | 0 | 0 | 0 | 4 |
| McPherson et al. 2004 | 5 | 5 | 0 | 0 | 0 | 0 |
| Thuiller et al. 2004 | 1 | 0 | 0 | 0 | 0 | 1 |
| Thuiller et al. 2003 | 2 | 0 | 0 | 0 | 0 | 2 |

497 **Appendix 3**. Full references for studies cited in Appendix 1 and 2.


498 Araújo, M. B. et al. 2005. Downscaling European species atlas distributions to a finer
499 resolution: implications for conservation planning. - Global Ecol. Biogeogr. 14: 17-
500 30.

501 Beale, C. M. et al. 2008. Opening the climate envelope reveals no macroscale associations
502 with climate in European birds. - Proc. Natl. Acad. Sci. USA 105: 14908-14912.

503 Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from
504 occurrence data. - Ecography 29: 129-151.

505 Freedman, A.H. et al. 2009. Modeling the effects of anthropogenic habitat change on savanna
506 snake invasions into African rainforest. - Conserv. Biol. 23: 81-92

507 Guisan, A. and Hofer, U. 2003. Predicting reptile distributions at the mesoscale: relation to
508 climate and topography. - J. Biogeogr. 30: 1233-1243.

509 Heikkinen, R. K. et al. 2007. Biotic interactions improve prediction of boreal bird
510 distributions at macro-scales. - Global Ecol. Biogeogr. 16: 754-763.

511 Huntley, B. et al. 2004. The performance of models relating species geographical
512 distributions to climate is independent of trophic level. - Ecol. Lett. 7: 417-426.

513 Huntley, B. et al. 2006. Potential impacts of climatic change upon geographical distributions
514 of birds. - Ibis 148: 8-28.

515 Huntley, B. et al. 2008. Potential impacts of climatic change on European breeding birds. -
516 PLoS ONE 3: e1439.

517 Luoto, M. et al. 2005. Uncertainty of bioclimate envelope models based on the geographical
518 distribution of species. - Global Ecol. Biogeogr. 14: 575-584.

519 McPherson, J. M. and Jetz, W. 2007. Effects of species' ecology on the accuracy of
520 distribution models. - Ecography 30: 135-151.

McPherson, J. M. et al. 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? - J. Appl. Ecol. 41: 811-823.

Parra, J. L. and Monahan, W. B. 2008. Variability in 20th century climate change reconstructions and its consequences for predicting geographic responses of California mammals. - Global Change Biol. 14: 2215-2231.

Pearson, R. G. et al. 2006. Model-based uncertainty in species range prediction. - J. Biogeogr. 33: 1704-1711.

Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. - Ecol. Model. 190: 231-259.

Thuiller, W. 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. - Global Change Biol. 9: 1353-1362.

Thuiller, W. et al. 2003. Generalized models vs. classification tree analysis: Predicting spatial distributions of plant species at different scales. - J. Veg. Sci. 14: 669-680.

Thuiller, W. et al. 2004. Effects of restricting environmental range of data to project current and future species distributions. - Ecography 27: 165-172.

Thuiller, W. et al. 2006. Vulnerability of African mammals to anthropogenic climate change under conservative land transformation assumptions. - Global Change Biol. 12: 424-440.

Venier, L. A. et al. 2004. Climate and satellite-derived land cover for predicting breeding bird distribution in the Great Lakes Basin. - J. Biogeogr. 31: 315-331.

542 **Appendix 4**. Attributes of studies that contained dispersal distances for species in our dataset.

| Study | Taxonomic Group | Details |
|---|---|---|
| Bowman et al. 2002 | mammals | Conducted a literature review to find maximum distance moved by adult mammals after translocation. |
| Cain et al. 1998 | plants | Measured dispersal distance for *Asarum canadense* via direct observations of seed movement by ants; searched the literature for measured dispersal distances for other woodland herbs. Some of these were directly observed and others were based on measured fall rates of seeds combined with typical wind speeds. |
| Paradis et al. 1998 | birds | Used survey data from the ringing scheme of the British Trust for Ornithology 1909-1994. Included only birds ringed and recovered during the breeding season (i.e. excluded migration distances). Estimated both natal and breeding dispersal distances. |
| Schneider 2003 | butterflies | Compiled mean distances reported in mark-release-recapture studies. |
| Smith and Green 2005 | amphibians | Compiled a list of the longest distances moved in both mark-recapture and displacement studies. |
| Sutherland et al. 2000 | mammals and birds | Compiled data on natal dispersal distances from a literature search. Most data were based on incidental observations. Did not accept data from "likely migrants". |

543 **Appendix 5**. Full references for studies cited in Appendix 4.

544 Bowman, J. et al. 2002. Dispersal distance of mammals is proportional to home range size. -
545 Ecology 83: 2049-2055.
546 Cain, M. L. et al. 1998. Seed dispersal and the Holocene migration of woodland herbs. - Ecol.
547 Monogr. 68: 325-347.
548 Paradis, E. et al. 1998. Patterns of natal and breeding dispersal in birds. - J. Animal Ecol. 67:
549 518-536.
550 Schneider, C. 2003. The influence of spatial scale on quantifying insect dispersal: an analysis
551 of butterfly data. - Ecol. Entomol. 28: 252-256.
552 Smith, M. A. and Green, D. M. 2005. Dispersal and the metapopulation paradigm in
553 amphibian ecology and conservation: are all amphibian populations metapopulations?
554 - Ecography 28**:** 110-128.
555 Sutherland, G. D. et al. 2000. Scaling of natal dispersal distances in terrestrial birds and
556 mammals. - Conserv. Ecol. 4: 44.
557
558
559

560 **Appendix 6**. Collinearity (Spearman r coefficients) between all continuous covariates

561 (n=4317).

562

| | latitude | area | resolution |
|---|---|---|---|
| area | -0.903 | | |
| resolution | -0.589 | 0.503 | |
| variables | 0.277 | -0.290 | -0.479 |

**Appendix 7.** Analysis of deviance table for the relationship between model accuracy, covariates and taxonomic group when studies that contributed more than half of the total number of species in one taxonomic group were removed. Presented are the differences in AIC and deviance between full and reduced models as well as the associated *p* value. The difference in degrees of freedom between full and reduced models was four for all comparisons and subsets. The full model includes number of variables, log(spatial extent), resolution and model type. When Huntley et al. 2006 is removed there are 2860 species from nineteen published studies, without Araújo et al. 2005 there are 2539 species from nineteen published studies and without Luoto et al. 2005 there are 4238 species from nineteen published studies.

574

| Studies subset | Model for comparison | Model terms | ΔAIC | ΔDeviance | *p* |
|---|---|---|---|---|---|
| Without Huntley et al. 2006 | Just intercept | +Taxonomic group | 33.46 | 41.46 | <0.0001 |
| | Full model | +Taxonomic group | 31.88 | 39.88 | <0.0001 |
| Without Araújo et al. 2005 | Just intercept | +Taxonomic group | 14.96 | 22.96 | 0.000129 |
| | Full model* | + Taxonomic group | 15.18 | 23.18 | 0.000117 |
| Without Luoto et al. 2005 | Just intercept | +Taxonomic group | 41.6 | 49.6 | <0.0001 |
| | Full model | +Taxonomic group | 41.76 | 49.76 | <0.0001 |

575

576 * Only includes spatial extent and resolution, none of the other covariates in the model
577 led to estimation

578